CHAPTER TEN

Adapting CRISPR/Cas9 for Functional Genomics Screens

Abba Malina*, Alexandra Katigbak*, Regina Cencic*, Rayelle Itoua Maïga*, Francis Robert*, Hisashi Miura*, Jerry Pelletier*,†,‡,1

*Department of Biochemistry, McGill University, Montreal, Quebec, Canada
†Department of Oncology, McGill University, Montreal, Quebec, Canada
‡The Rosalind and Morris Goodman Cancer Research Center, McGill University, Montreal, Quebec, Canada
1Corresponding author: e-mail address: jerry.pelletier@mcgill.ca

Contents

Abstract

The use of CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein) for targeted genome editing has been widely adopted and is considered a "game changing" technology. The ease and rapidity by which this approach can be used to modify endogenous loci in a wide spectrum of cell types and organisms makes it a powerful tool for customizable genetic modifications as well as for large-scale functional genomics. The development of retrovirus-based expression platforms to simultaneously deliver the Cas9 nuclease and single guide (sg) RNAs provides unique opportunities by which to ensure stable and reproducible expression of the editing tools and a broad cell targeting spectrum, while remaining compatible with *in vivo* genetic screens. Here, we describe methods and highlight considerations for designing and generating sgRNA libraries in all-in-one retroviral vectors for such applications.

Methods in Enzymology, Volume 546
ISSN 0076-6879
http://dx.doi.org/10.1016/B978-0-12-801185-0.00010-6

1. INTRODUCTION

The rise of functional genetic screens in mammalian cells and animal models owes a considerable debt to RNA interference (RNAi) technology. RNAi allows for broad, systemic, and unbiased inquiry into complex biological systems in a wide variety of contexts due in large part to development of genome-wide multiplexed pooled short-hairpin RNA (shRNA) library-based screening methods. Yet despite its proven track record, state-of-the-art RNAi-based screens have their drawbacks: (1) targets are limited to the exome; (2) a substantial portion of shRNAs often yield incomplete and unpredictable knockdown efficiencies, which can be insufficient to elicit the desired phenotype of interest; (3) many shRNAs have "off-target" effects that increase the number of spurious hits and lead to erroneous interpretations; and (4) although this can be partially mitigated by increasing the diversity and gene coverage of shRNA targets, it comes at the expense of increased library-pool size and assay complexity.

Applying modern genome editing tools to genetic screens aims to solve many of these problems. While modular transcription factor-based genome editing technologies such as zinc-finger and transcription activator-like effector-based nucleases (ZFNs and TALENs, respectively) have been demonstrated to be reliable and powerful on a one-by-one gene targeting basis, they are all but impractical to implement at a genome-wide scale due to their inherent bulky, pair-wise, and iterative design parameters. In contrast, CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein)-based genome editing has shown tremendous promise as a versatile and practical gene targeting technique that would be amenable to genetic screening approaches. Based on a bacterial adaptive immune response that targets invading foreign viral and plasmid DNA, the type II CRISPR system uses an RNA-guided DNA endonuclease (Cas9) to cleave DNA in a sequence-specific manner through a ~20 nt RNA–DNA base match (Jinek et al., 2012). Thus, Cas9 can be readily programmed to introduce double-stranded breaks in virtually any genomic locus through simple alteration of a ~20-bp cognate single guide RNA (sgRNA) when coexpressed in a cell (Cong et al., 2013; Jinek et al., 2013; Mali, Yang, et al., 2013). This inherent flexibility and design simplicity makes CRISPR/Cas9 genome editing easily adaptable for mammalian whole-genome screens and indeed we are beginning to see its application in such settings (Koike-Yusa, Li, Tan, Velasco-Herrera, & Yusa, 2013;

Shalem et al., 2014; Wang, Wei, Sabatini, & Lander, 2014; Zhou et al., 2014). Here, we present methodology and discuss issues pertaining to the use of CRISPR/Cas9 for positive-selection screens.

2. ALTERING THE VECTOR DESIGN FOR HIGH-THROUGHPUT SCREENS

The first step in any successful CRISPR/Cas9-based genetic screen is choosing the appropriate method of expression of the two key editing components, Cas9 and its cognate sgRNA. Two approaches dominate the literature: either expressing Cas9 and sgRNA from separate vectors or expressing both in an "all-in-one" vector design. While there are some advantages to independently expressing each part (e.g. the ability to use two different selection markers), we opt for simultaneous delivery given the convenience of a linked single-vector format and, importantly, a more consistent level of expression of either Cas9 and sgRNA not only in terms of selectability but also in terms of stoichiometry, the latter of which has been shown to be important for mitigating off-target cleavage events (Hsu et al., 2013; Pattanayak et al., 2013). Retroviral plasmids provide a convenient way for achieving this given their broad tropism, adjustable levels of infections and expression, and the ability to enrich for permanent and successful integration with a selectable marker (either fluorescence or drug resistance). While we have previously reported the construction and characterization of "all-in-one" retroviral-based vectors coexpressing sgRNAs and Cas9 (from the murine U6 small nucleolar RNA promoter and from the SV40 or Spleen focus-forming virus (SFFV) promoters, respectively) (Malina et al., 2013), we have since modified their design to better suit high-throughput screening purposes by engineering unique restriction sites 17 nucleotides upstream of the U6 transcription start site and at the junction of the crRNA/tracrRNA fusion, in order to facilitate the insertion of oligonucleotides harboring guide sequences, streamlining the process for the generation of sgRNA-based libraries (Fig. 10.1A and B). To distinguish these vectors from our first generation pQCiG and pLC series, we refer to them as pQCiG2 and pLCiG2. These vectors, like their predecessors, express human codon-optimized Cas9 from *Streptococcus pyogenes* (SpCas9) with a 3xFlag epitope tag and two NLS (nuclear localization signal) tags at the N-terminus. Cas9 expression can be explicitly monitored given that its transcription is linked to GFP via an EMCV IRES (Fig. 10.1A and B). We made certain that these subtle changes in sequence would not interfere with

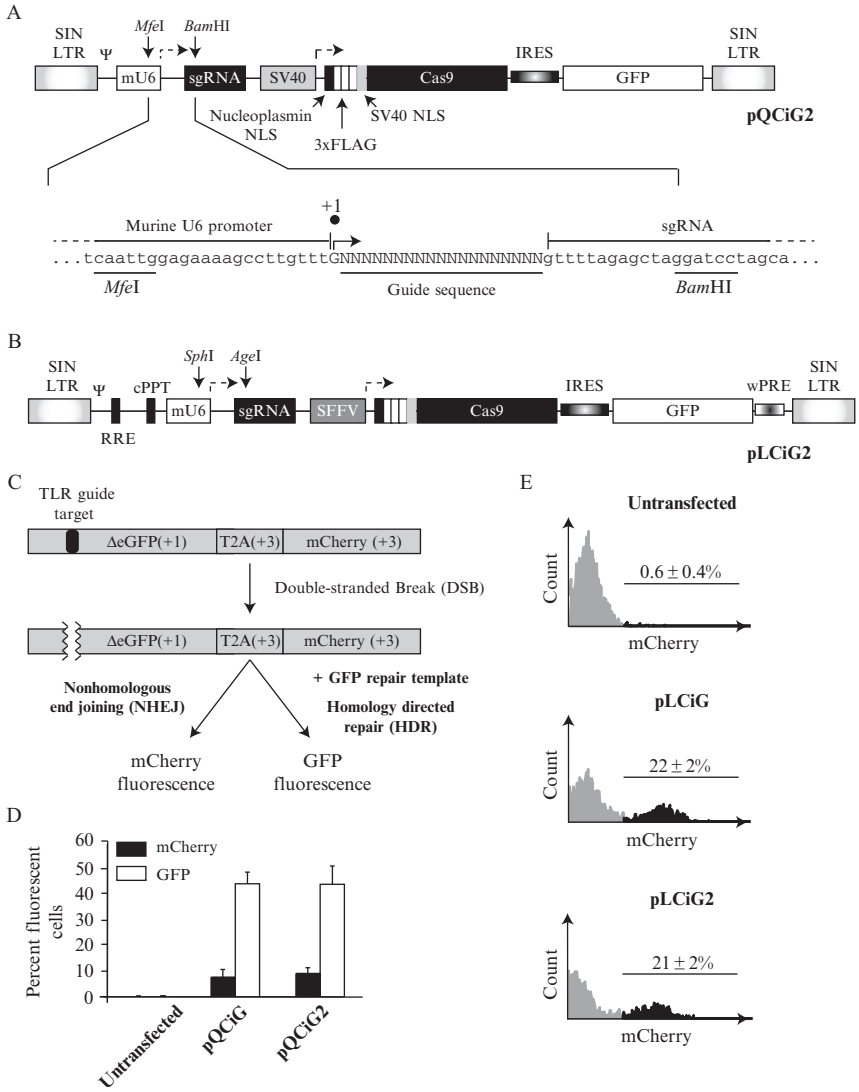

Figure 10.1 Design of retrovirus vectors for codelivery of Cas9 and sgRNAs that are compatible with large-scale guide library generation. (A) Schematic diagram of pQCiG2-based vector driving expression of Cas9, GFP, and sgRNAs. The unique *Mfe*I and *Bam*HI sites are indicated and are present within the murine U6 promoter and sgRNA, respectively. Right-angled arrows denote the site of transcription initiation. The expanded view illustrates the nucleotide sequence spanning a portion of the mU6 promoter, the start of transcription (G at +1), the 19 nt guide sequence, and the 5′ end of the sgRNA. (B) Schematic diagram of pLCiG2-based lentivirus vector driving expression of Cas9, GFP, and sgRNAs. The unique *Sph*I and *Age*I sites are indicated

Cas9-driven genome editing by assaying the relative cleavage efficiencies of the new version compared to the old version using the "traffic light reporter" (TLR) system, an assay that simultaneously measures the frequency of NHEJ and HDR following a Cas9-induced DSB (Certo et al., 2011). Both versions of the Cas9/sgRNA retroviral vectors stimulated NHEJ to similar extents in 293T cells, indicating that the introduced changes had not impaired editing activity (Figs. 10.1D and E).

Our original sgRNA design incorporated elements from published work by Church and coworkers (Mali, Yang, et al., 2013; Fig. 10.2A, top design), but more recent publications have used a significantly altered sgRNA layout. Two new versions are notable: (1) a version termed sgRNA.2.1, which extends the crRNA:tracrRNA scaffold by four nucleotides, has been reported to improve cleavage efficiency but with concomitant decrease in on-target versus off-target specificity (Pattanayak et al., 2013) and (2) a version incorporating the aforementioned extension and which also mutates a U-rich stretch immediately downstream of the guide sequence, which has been suggested to function as an RNA Pol III transcription termination signal (Fig. 10.2A, sgRNA.3). This has been reported to reduce nucleolar localization of Cas9 (Chen et al., 2013). We evaluated whether these

and cleave within the murine U6 promoter and sgRNA, respectively. (C) A schematic of the traffic light reporter (TLR) assay (Certo et al., 2011). The sgRNA guide target sequence is engineered in the GFP open reading frame (ORF) and shifts the reading frame leading to premature translation termination. The GFP ORF (+1 frame) is fused out of frame to the T2A ribosome "skipping" sequence (Szymczak-Workman, Vignali, & Vignali, 2012) and the mCherry ORF (+3 frame). Induction of a DSB at the guide target sequence will result in mutagenic repair by NHEJ which, in one of three cases, will place the disabled GFP ORF in-frame with mCherry, yielding $mCherry^+$ cells. Exogenously supplying a truncated GFP donor plasmid *in trans* will result in GFP fluorescence as a result of HDR of the TLR GFP ORF. Since our vectors express GFP as a reporter, we could not score for HDR activity but rather used the percentage of GFP^+ cells as an assessment of transfection efficiency and mCherry fluorescence as a gauge of relative NHEJ repair efficiency. Both vectors harbored a previously described guide sequence (TLR: $^{5\prime}$GAGCAGCGTCTTCGAGAGTG$^{3\prime}$) that targets a unique site embedded within the GFP ORF of the TLR (C) (Malina et al., 2013). (D) Assessment of pQCiG- and pQCiG2-mediated NHEJ in a stably integrated TLR reporter 293T cell line with a stably integrated TLR reporter locus. Cells were transfected with pQCiG or pQCiG2 (1.5–3 μg) and analyzed by flow cytometry 6 days later. GFP fluorescence measures transfection efficiency whereas mCherry fluorescence scores for NHEJ repair events. $n = 3$, error bars represent SEM. (E) Assessment of pLCiG- and pLCiG2-mediated NHEJ in a stably integrated TLR reporter 293T cell line. Cells were transfected with pLCiG or pLCiG2 and analyzed by flow cytometry 6 days later. Shown is a representative histogram illustrating the percent of $mCherry^+$ cells and the mean percent fluorescence. $n = 4$, error is SEM.

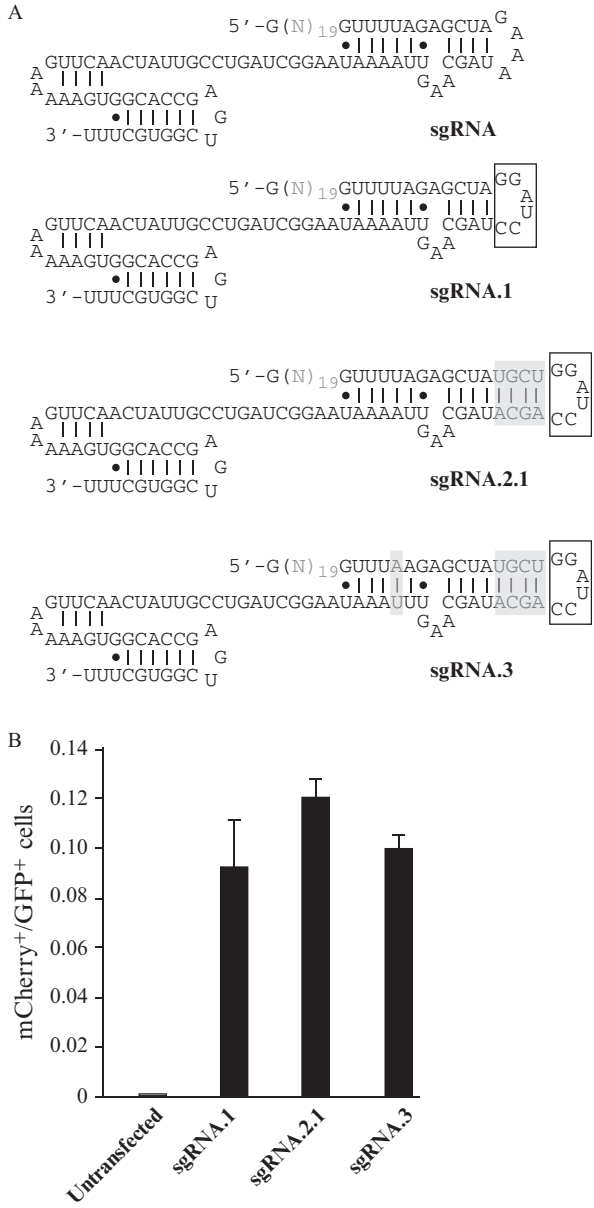

Figure 10.2 Assessment of NHEJ repair efficiency mediated by different sgRNA variants. (A) Predicted secondary structure (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) of chimeric RNAs showing the first guanine arising from the transcription initiation site followed by the guide region $(N)_{19}$, for four different sgRNAs. The open box denotes the crRNA/tracrRNA junction where a *Bam*HI site was inserted to generate sgRNA.1 (*Age*I

changes would produce any significant functional differences, but could not detect any differences in gene editing efficiencies among the three sgRNAs (Fig. 10.2B). In light of this and given our preference to maintain a high ratio of on-target to off-target specificity, our retroviral vectors retain the original sgRNA.1 configuration.

3. CONSTRUCTION OF sgRNA LIBRARIES

3.1. Guide sequence prediction

Prediction of guide sequences can be accomplished by manually inspecting annotated gene sequences (when only a small number of guides is required) or by using one of several design tools available at the time of this writing (Table 10.1). In either case, one first locates a sequence of interest bearing a protospacer-adjacent motif (PAM), which is essential for recognition by Cas9. Our vectors use a humanized version of Cas9 protein that originates from *S. pyogenes* and is the one most frequently used in the literature, owing to its short PAM target sequence ($^{5'}$NGG$^{3'}$) and thus high prevalence in the genome (Jiang, Bikard, Cox, Zhang, & Marraffini, 2013; Jinek et al., 2012). Although it has been reported that $^{5'}$NAG$^{3'}$ can also be used as a PAM by *S. pyogenes* Cas9, it is much less efficiently recognized (Jiang et al., 2013), and probably very rarely so at limiting Cas9 cellular concentrations (Wu et al., 2014), thus we generally do not consider it when designing guide sequences. After locating a PAM sequence, the adjacent 20 upstream nucleotides to the PAM are chosen as the guide sequence. Should the 20th nucleotide not end with guanosine, we forcibly terminate the sequence with a 5′ guanosine, which is a necessary requirement for U6 transcription initiation but has little effect on the rate of target cleavage even when unmatched (Fu, Sander, Reyon, Cascio, & Joung, 2014; see also Fig 10.1A). Mismatches between the PAM proximal region of the target and the sgRNA are known to more adversely affect Cas9 endonuclease activity (Fu et al., 2013; Hsu et al., 2013; Jinek et al., 2012; Mali, Aach, et al., 2013;

in the case of pLCiG2). The grey shaded areas denote sequence differences between sgRNA.1, sgRNA2.1, and sgRNA.3. Note that our sgRNA.2.1 and sgRNA.3 designs differ from the originals in harboring a *Bam*HI site at the crRNA/tracrRNA junction. (B) Assessment of normalized NHEJ repair efficiency in a stably integrated TLR reporter 293T cell line. Cells were transfected with pQCiG2 (1.5–3 μg) expressing the indicated sgRNAs and analyzed by flow cytometry 6 days later. GFP fluorescence was used to track transfection efficiency, whereas mCherry was used to monitor NHEJ. GFP values (transfection efficiency) ranged from 32% to 51%. $n = 3$, error bars represent SEM.

Table 10.1 CRISPR/Cas9 guide design tools

Tool name	Web interface	Target genomes[a]	URL	Off-target analysis[b]
CRISPR design	Yes	15	http://crispr.mit.edu http://www.broadinstitute.org/mpg/crispr_design/	Yes
E-CRISP	Yes	18	http://www.e-crisp.org/E-CRISP/designcrispr.html	Yes
Cas9 design	Yes	7	http://cas9.cbi.pku.edu.cn/index.jsp	No
CasOT	No	Any	http://eendb.zfgenetics.org/casot/index.php	Yes
CRISPR sgRNA design tool	Yes	3	https://www.dna20.com/eCommerce/cas9/input	No
CasFinder	No	Any	http://arep.med.harvard.edu/CasFinder/	Yes
flyCRISPR	Yes	Fly	http://flycrispr.molbio.wisc.edu/tools	Yes
DRSC CRISPR finder	Yes	Fly	http://www.flyrnai.org/crispr/	Yes
ZiFiT Targeter	Yes	Any	http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx	No
CRISPy	Yes	CHO	http://staff.biosustain.dtu.dk/laeb/crispy/	Yes[c]
GT-Scan	Yes	32	http://gt-scan.braembl.org.au/gt-scan/submit	Yes
CHOPCHOP	Yes	9	https://chopchop.rc.fas.harvard.edu/	Yes[d,e]

[a]Refers to the number or nature of species that the software allows one to analyze.
[b]Refers to whether the software is capable of predicting off-target sites based on sequence similarity and location adjacent to a PAM.
[c]Can only scan for sites matching 13 nucleotides + NGG.
[d]Can scan alternate Cas9 PAM motifs.
[e]Can also output flanking primer sequences to, and identify restriction sites within, the target site.

Pattanayak et al., 2013), which lends support to the notion that an 8–12 nts "seed" sequence upstream of the PAM drives Cas9-mediated cleavage efficiencies (Jinek et al., 2012; Semenova et al., 2011). In order to minimize potential off-target cleavage sites, and given the more stringent requirement

for homology between the "seed" region and the sgRNA, we typically heuristically align only the first 12 nucleotides of the chosen sequence plus all four iterations of the PAM to annotated online genome databases, with sequences that result in the least number of perfect matches being preferred. More recent genome-wide ChIP-seq-based analyses have suggested a far greater tolerance for mismatches driving Cas9 DNA binding, with enriched genomic regions being frequently characterized by "seed" sequences that can be as short as five nucleotides, although most of these sites were only rarely altered when sequenced directly (Kuscu, Arslan, Singh, Thorpe, & Adli, 2014; Wu et al., 2014). Nevertheless, as a precaution, we recommend designing at least three sgRNAs for each locus to control for potential off-target effects. Further points in sgRNA design that should also be considered:

1. When targeting genes encoding mRNAs, sgRNAs targeting the last coding exon have been reported to be less effective than those targeting earlier exons (Wang et al., 2014). As well, we would recommend that users avoid targeting the region that harbors the first AUG codon since genes may have in-frame downstream AUG (and even non-AUG) initiation codons that can be used and give rise to functional truncated products (Ellison & Bishop, 1996). Rather it is probably a safer bet to target somewhere in the middle of a gene when disruption of function is desired.
2. It has been reported that sgRNAs that target the transcribed strand are less effective than those targeting the nontranscribed strand (Wang et al., 2014).
3. Be aware that the sgRNAs are transcribed by RNA Polymerase III, whose termination signal is a stretch of four or more sequential Us (Nielsen, Yuzenkova, & Zenkin, 2013; Orioli et al., 2011) and guide sequences that are U-rich have been shown to decrease sgRNA abundance (Wu et al., 2014). Therefore, avoid guides that have stretches of three or more Us.
4. Recent crystal structure data of sgRNA-bound Cas9 have revealed protein:RNA interactions between residues Arg71-G18 and Arg447-U16, which correspond to the 3rd and 5th residue upstream from the crRNA:tracrRNA scaffold (Nishimasu et al., 2014). This is in line with other recently reported data that high-performing sgRNAs displayed a preference for four purines adjacent to the PAM (Wang et al., 2014), so it might be worthwhile to prioritize guides with a G residue three nucleotides upstream of the PAM, if one has that option when choosing.

5. sgRNAs with very high or very low-GC content should be avoided (Wang et al., 2014).
6. Finally, ensure that your guide sequences are absent of restriction sites used for cloning (*Mfe*I/*Bam*HI or *Sph*I/*Age*I).

3.2. Cloning of guide templates

The construction of guide libraries uses either pools of oligonucleotides derived from small-scale synthesis or from highly parallel approaches (Fig. 10.3).

3.2.1 Layout of the guide template

The template for cloning into pQCiG2 is: $^{5'}$CAATTG-GAGAAAAGCCTTGTTTG(N)$_{19}$ GTTTTAGAGCTAGGATCCTAGC$^{3'}$ (where the *Mfe*I and *BamH*I sites are underlined and 19 nucleotide guide region represented by N). For pLCiG2, the template is: $^{5'}$GCATGC-GAGAAAAGCCTTGTTTG(N)$_{19}$ GTTTTAGAGCTAACCGGTTAGC$^{3'}$ (where the *Sph*I and *Age*I sites are underlined).

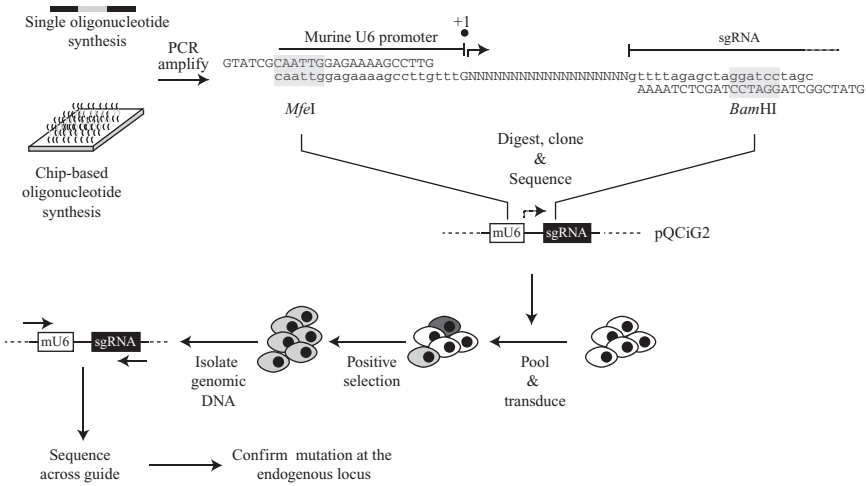

Figure 10.3 Schematic representation of sgRNA library generation and pooled screening strategy. Oligonucleotides are individually synthesized or *en masse* on a microarray chip. These are then PCR amplified to incorporate vector compatible restriction sites. The sequence of the primers and template shown are compatible with cloning into pQCiG2 (see text for details for cloning into pLCiG2). Library pools of guides are then used for screening purposes. Following isolation of genomic DNA from positively selected cells, amplification by PCR across the guide region is performed and the guide identified by sequencing. Modification at the expected locus is then confirmed using the T7 endonuclease I assay, SURVEYOR assay, or sequencing of PCR products.

3.2.2 Initial guide library preparation

Depending on the size and complexity of the required sgRNA library, one pools oligos ordered either individually prealiquoted in 96- or 384-well dishes (e.g., IDT, Coralville, IA), or synthesized *en masse* on a chip as an array and liberated following acid hydrolysis (e.g., Oligomix from LC Sciences Inc., Houston, TX). In our experience, we get much lower rates of mutant clones when derived from pools of individually synthesized oligonucleotides than those from arrays (~80–90% vs. 30–50% produce error-free clones, respectively).

3.2.3 PCR amplification of pooled oligonucleotide templates

If ordered on an individual basis, oligonucleotides should first be pooled at an equimolar ratio and then amplified using Forward ($^{5'}$GTATCGCAATTGGAGAAAAGCCTTG$^{3'}$ for pQCiG2 and $^{5'}$GTATCGGCATGCGAGAAAAGCCTTG$^{3'}$ for pLCiG2) and Reverse ($^{5'}$GTATCGGCTAGGATCCAGCTCTAAAA$^{3'}$ for pQCiG2 and $^{5'}$GTATCGGCTAACCGGTTAGCTCTAAAA$^{3'}$ for pLCiG2) primers (Fig. 10.3). PCR conditions are as follows:

Reagent amounts

- 5 μl 10 × ThermoPol buffer with $MgCl_2$
- 2.5 μl Forward Primer (10 μ*M*)
- 2.5 μl Reverse Primer (10 μ*M*)
- 1 μl of Oligonucleotide Template (100 ng/μl)
- 1 μl dNTPs (10 m*M*)
- 0.25 μl Vent DNA Polymerase (NEB) (2 U/μl)
- 37.75 μl dH_2O

Thermocycler reaction conditions

- 94 °C for 3 min (Initial denaturation)
- 30 cycles of 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 1 min
- 72 °C for 10 min (Final extension)

The PCR conditions and reagents are slightly different if amplifying oligos from arrays, and are as follows:

Reagent amounts

- 10 μl 5 × Phusion buffer
- 1 μl Forward Primer (20 μ*M*)
- 1 μl Reverse Primer (20 μ*M*)
- 1 μl Oligo Template (0.5 ng/μl)
- 1 μl dNTPs (10 m*M*)
- 0.5 μl Phusion High-Fidelity DNA polymerase (NEB) (2 U/μl)

1 μl 30% DMSO
34.5 μl dH_2O
Thermocycler reaction conditions
98 °C for 30 s (Initial denaturation)
30 cycles of 98 °C for 10 s, 54 °C for 30 s, and 72 °C for 25 s
72 °C for 5 min (Final extension)
Confirm amplification of the desired PCR products by analyzing an aliquot (5 μl) on a 2% agarose gel to verify the presence of a single band (76 bp).

3.2.4 Digestion and ligation of the guides into vector backbone

The PCR product is purified using a PCR Purification Kit (e.g., QIAquick kits [Qiagen] or EZ-10 Spin Column PCR Products Purification Kit [Bio Basic Inc.]) following the manufacturer's recommendations and the eluent digested with *Mfe*I/BamHI-HF or *Sph*I/*Age*I (NEB) depending on the desired target vector. Ligations into the appropriate vector are performed according to standard techniques (Green & Sambrook, 2012). Make sure to include a "vector-only" ligation control. Following ligations, 2 μg of glycogen is added to each ligation and the volume is increased to 100 μl with ddH_2O, followed by two consecutive ethanol precipitations and 70% ethanol washes. The precipitate is resuspended in 20 μl ddH_2O and is ready for transformation by electroporation.

3.2.5 Assessing ligation efficiency

An aliquot (1 μl) of the ligation is used for a test chemical transformation and the ratio of colonies from the "vector + insert" ligation reaction to "vector-only" reaction is determined. We proceed with large-scale transformation if we obtain at least a 10:1 ratio of colonies in the "vector + insert" plate relative to the "vector-only" plate. We also process at least 24 minipreps and have them sequenced to assess the quality of the clones and representation of the library.

3.2.6 Large-scale transformation of the guide library

To generate the guide RNA-expressing retroviral plasmid library as bacterial clones, we use Electromax-competent DH10B cells for pQCiG2 vectors or Electromax-competent Stbl4 for pLCiG2 (Life Technologies) and a Bio-Rad Gene Pulser using the following conditions of 2.0 kV, 200 Ω, and 25 μF. We follow the manufacturer's recommendation using 1 μl of ligation/100 μl of competent cells. From 1 ml of culture following transformation, aliquots of 1, 2, and 10 μl are taken and plated onto LB + 100 μg/ml

carbenicillin plates to assess the efficiency of transformation. The remaining culture is kept at 4 °C overnight. Once transformation efficiency has been determined, the reserved culture material is plated onto large square LB plates (245 mm × 245 mm) containing 100 μg/ml carbenicillin to obtain 1000–2000 (if colonies are to be individually picked) or 10,000 (if colonies are to be pooled) colonies/plate. We prefer the use of carbenicillin over ampicillin since it is more stable than the latter and results in fewer satellite colonies during library generation.

3.2.7 Checking the quality of the guide library

Ninety-six colonies are seeded into deep-well 96-well plates (VWR® 96-Well Deep Well Plates; cat. no.: 82006–448) containing 1.5 ml of Terrific broth (TB) + 100 μg/ml carbenicillin. The plates are sealed with an air-pore sheet (Qiagen cat. no.: 19571). Following growth in a 37 °C shaker for 24 h, we isolate plasmid DNA using a QIAprep 96 Turbo Miniprep Kit (QIAgen), which is then submitted for sequencing.

3.2.8 Bulk harvesting of bacterial-transformed guide library

For some applications, it may be sufficient to harvest the plated colonies in bulk and use the resulting pool directly in a screen. This is achieved by pipetting 50 ml of TB + 100 μg/ml carbenicillin directly onto each plate and using a flat rubber policeman to gently scrape the colonies off the plate into a sterile 2-l flask. The nature of the screen will determine the desired library complexity, but we use 500 ml of TB + 100 μg/ml carbenicillin per pool aiming for complexities of 10,000–20,000 clones/pool. After growth at 37 °C for ~6 h, the plasmid DNA can be isolated using standard procedures (Green & Sambrook, 2012) or a commercial maxiprep kit (e.g., Plasmid Maxi Kit; Qiagen).

3.2.9 Arraying individual bacterial guide library clones

Although significantly more expensive and labor intensive than a nonarrayed library, our preference is to generate arrayed, sequence-verified libraries since these are renewable resources with greater flexibility. Here, individual colonies are picked into deep-well 96-well plates containing 1.5 ml of TB + 100 μg/ml carbenicillin and covered with an air-pore sheet (Qiagen cat. no.: 19571). Following growth for 24 h at 37 °C, 50 μl aliquots are transferred to two 96-well plates (Falcon cat. no.: 353910) containing 50 μl TB + 100 μg/ml carbenicillin + 50% glycerol, sealed and stored at −70 °C as Master plates. The remainder of the culture is processed to

prepare miniprep DNA that is then used for sequencing across the sgRNA insert. Alternatively, colonies can be picked into 384-well plates containing 65 μl of TB + 100 μg/ml carbenicillin. Following growth overnight at 37 °C, 2 μl of a 1/10th dilution of the culture is directly used in a PCR as template for amplification across the guide sequence. The PCR product is purified using Agencourt AMPure XP—PCR Purification (Beckman-Coulter) and used directly for sequencing. The position and identity of each clone in the Master Plates is recorded.

Once arrayed, the library pools are made by identifying the coordinates of the clones of interest and then thawing the plates at room temperature. The plates are then briefly centrifuged, a small hole is made by piercing through the aluminium foil cover and 1 μl of the desired bacterial culture corresponding to the clone of interest is removed. The puncture hole is sealed using a small aluminium foil patch. This method avoids potential well cross-contamination due to aerosol generation that could arise if the entire cover was removed. The 1 μl aliquot is used to seed 1 ml of TB + 100 μg/ml carbenicillin and grown at 37 °C to saturation (~24 h). The following day, the individual bacterial cultures are pooled into a 2-l flask containing 500 ml TB + 100 μg/ml carbenicillin, grown for 6 h and processed for plasmid DNA isolation.

4. RETROVIRAL TRANSDUCTION OF THE GUIDE LIBRARY

The resulting library is then used for virus preparation using standard techniques (Barde, Salmon, & Trono, 2001; Swift, Lorens, Achacoso, & Nolan, 2001). Depending on the viral vector backbone, either helper-free stable virus producing Phoenix cell line is used (in the case of pQCXiG2-based libraries), or 293T/17 (ATCC) cells are cotransfected with packaging and VSV-G envelope vectors (in the case of pLCiG2-based libraries). One advantage of using pseudotyped lentivirus is the ability to generate large quantities of library-pool transducing viral supernatant preps which can then be concentrated, titered, aliquoted, and frozen for later use (for further details, see Kutner, Zhang, & Reiser, 2009). The viral MOI (multiplicity of infection) is determined by serial dilution of the preparation on 293T cells and through measurement of the fraction of GFP expressing cells as determined by FACS (we aim to get to at least 5–10% GFP^+ cells, which is in the linear range of viral transduction). The amount of cells plated for viral preparation will depend on the desired library complexity: generally speaking, we want to make enough virus to infect cells at an MOI of ~0.2–0.1 (which

ensures that only one sgRNA is expressed per cell for the vast majority of the population) with also at least 1000 infected cells/construct (which maintains the library complexity—see notes below).

5. NOTES ON SCREENING DESIGN PARAMETERS

Each genetic screen will entail designs unique to the respective experiment. Rather than present specifics about one particular screen, it is more practical to consider general attributes that will impact on most screens:

1. *The nature of the phenotype and the strength of the selective pressure*. One of the most important determinants of the success of a screen is how well the desired phenotype can be distinguished from baseline. It should be robust and with little variation. Positive-selection screens looking for cooperating tumor suppressors or lesions that impart drug resistance embody these features. The time to phenotype onset will dictate the duration of the experiment and the strength of the selective pressure. Greater selective pressure enhances the phenotypic shifts in sgRNA representation under shorter periods of time, but can also lead to increased variability among replicates and can lead to a loss of representation of sgRNA species (especially those at the lower end of abundance) due to a sudden population bottleneck. This can be partially mitigated by increasing the number of infected cells per construct. While the optimal amount of selective pressure will have to be determined empirically in pilot screens (ideally with help of positive controls), we generally strive for ~25% loss of cell population following a given toxic treatment, which balances good reproducibility, selective pressure, and maintenance of sgRNA construct abundance.
2. *Maintaining library complexity during propagation of cells*. Many factors will determine the appropriate pool size and consequent sgRNA library representation in a cell population over the course of the screening process (e.g., cell line infectability, number of replicates, rate of allele modification, etc.). Following a successful screen, we generally will infer the representation of sgRNAs through the use of next-generation sequencing. The number of spurious reads (or baseline noise level) that arise from a massive parallel sequencer is typically in the range of 50–100 counts, and therefore as a rule of thumb we typically try to ensure that at least 1000 cells/construct are infected at the onset (which should result on average in roughly a ~10–20-fold increase in sgRNA read counts above baseline noise). Moreover, if over the course of a screen the cells need to be split,

it is critical to ensure that at each split the full library representation that was initially used at the start of the experiment will be maintained. If too many cells are removed during propagation, the representation of the library becomes skewed.

3. *The availability of positive and negative controls*. Although one does not always have access to positive controls when undertaking novel screens, their availability will significantly facilitate assay development and optimization. Pilot screens testing a series of serial dilutions of the positive control can be used to tease out the limits of detection for a given sgRNA and inform on the required library complexity. As well, we make sure to include multiple negative controls, both "scrambled" sgRNAs that do not match to any region in the genome as well sgRNAs that are known to cleave genes or loci that when disrupted are neutral for most phenotypes (e.g., AAVS1 for human cells or the ROSA26 locus for mouse). These are vital to score for the relative increase or decrease in sgRNA output following a successful screen.
4. *Tracking each step*. Our vectors harbor a GFP marker (which is neutral in most settings) allowing us to document infection efficiencies throughout the experiment.
5. Is monoallelic, biallelic, or multiallelic (in the case of pseudodiploid cells) modification required for the phenotype of interest? The efficiency of locus modification by CRISPR/Cas9 in high-throughput screens has been reported to range from 13% to >90% (Koike-Yusa et al., 2013; Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014). Although the reasons for this variation are unclear, it could relate to differences in guide targeting efficiency, MOI, cell line, the ratio of Cas9:sgRNA cellular levels, and the methods of library delivery. Given these potential issues, it is important to try to understand the phenotype(s) that is expected and whether all alleles of the target need to be inactivated and how the delivery system chosen for the screen will impact on this.
6. Different guides to the same target should yield the same phenotype. If this is not the case, we recommend generating additional sgRNAs to resolve the discrepancy. A recent publication has indicated that guide sequences with 17 or 18 nucleotides complementarity (called "tru-gRNAs") show reduced mutagenesis at off-target sites without sacrificing on-target editing efficiencies (Fu et al., 2014) and this feature could easily be incorporated into guide library design.
7. Be aware that loss of a particular sgRNA can occur during virus generation, which can happen due to a given sgRNA-affecting viral

replication and/or packaging, or may simply be due to the inactivation of an essential host gene in the packaging cell line. Deep sequencing of the library pool before and after virus production will shed information on this and is recommended.

8. To date, four large-scale screens have been published using CRISPR/Cas9 and nonarrayed sgRNA libraries (Koike-Yusa et al., 2013; Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014) and there are several lessons to be learnt from these:
   - **A.** Two screens engineered their cell lines to constitutively express Cas9 (Koike-Yusa et al., 2013; Zhou et al., 2014), whereas a third engineered a doxycycline-inducible Cas9 in the line of interest (Wang et al., 2014). Zhang and colleagues performed negative and positive-selection screens with a delivery system similar to the one described above by us (Shalem et al., 2014). Developing cell lines that express Cas9 is more labor intensive and requires prescreening of cell clones to identify the ones with highest editing efficiency since this can vary between clones and may be a consequence of variations in Cas9 expression levels (Zhou et al., 2014). Furthermore, the clonal nature of the cell might influence the phenotypic outcome of particular screen rendering it less widely applicable.
   - **B.** Wei and colleagues (Zhou et al., 2014) also ectopically expressed OCT1, a transcription factor shown to boost U6 promoter activity (Lin & Natarajan, 2012) in their line of interest. This added feature may increase sgRNA expression and should be piloted to assess whether the gain in sgRNA levels obtained with higher OCT1 levels translates into higher mutation efficiency, which could also influence the measured phenotype.
   - **C.** RNAi was *not* universally successful in validating the sgRNAs identified from the screens. The ability to phenocopy the results obtained with sgRNAs tended to correlate with knockdown efficiency (Koike-Yusa et al., 2013; Shalem et al., 2014).
   - **D.** In one screen, complementation with cDNAs was successful at reverting the phenotype (Koike-Yusa et al., 2013) and may be a better approach at validating "hits" than using shRNAs, assuming that the mutant allele is not functioning in a dominant-negative or gain-of-function manner.
9. The recent description of CRISPR/Cas9 gene editing in mice both *ex vivo* (where CRISPR/Cas9 was expressed in cultured primary lymphoma cells via retroviral transduction and later reimplanted) and

in vivo (in which the CRISPR/Cas9 system was delivered via hydrodynamic injections to directly modify hepatocytes *in situ*), raises the exciting possibility of performing CRISPR-based sgRNA screens in a live mammalian model organism (Malina et al., 2013; Yin et al., 2014).

6. DECODING "HITS" FROM POSITIVE SELECTION SCREENS INVOLVING sgRNA LIBRARY POOLS

Once cells are obtained following a positive selection screen, we identify the guide sequence responsible for the phenotype by amplifying across the guide of the integrated retroviral-derived construct in the cells of interest. Genomic DNA from the clone(s) of interest is isolated using standard techniques (Green & Sambrook, 2012) and the guide region amplified by PCR. In our experience, the guide region can be amplified quite specifically.

Reagent amounts

5 μl 5 × Phusion Buffer

1 μl Primer Mix (10 μ*M* each; Trigger ID F: $^{5'}$AGCCCTTTGTACACCCTAAGCCTC$^{3'}$

Trigger ID R: $^{5'}$CTAACTGACACACATTCCACAGGG$^{3'}$)

0.5 μl dNTPs (10 m*M*)

1 μl Genomic DNA from pQCiG2 infected cells (100 ng/μl)

0.15 μl of Phusion High-Fidelity DNA polymerase (NEB) (2 U/μl)

17.35 μl ddH_2O

Thermocycler reaction conditions

98 °C for 30 s (Initial denaturation)

25 cycles of 98 °C for 10 s, 57 °C for 30 s, and 72 °C for 30 s

72 °C for 10 s (final extension)

The PCR product is then purified using a PCR Purification Kit (e.g., Qiaquick kits [Qiagen] or EZ-10 Spin Column PCR Products Purification Kit [Bio Basic Inc.]) following the manufacturer's recommendations and directly sequenced using the sequencing primer Psi: $^{5'}$AGCCCTTTGTACACCCTAAGC$^{3'}$. Once the guide sequence has been successfully identified as a potential "hit," we then confirm that the endogenous locus has been mutated using the original genomic preps and perform either a T7 endonuclease I assay (Reyon et al., 2012) or SURVEYOR assay (Transgenomic), or, if a more thorough examination of the kinds of sequence alterations is desired, through sequencing on an Ion Torrent personal genome machine (Malina et al., 2013).

7. CONCLUSION

CRISPR/Cas9 has much to offer in complementing RNAi-based screens. The larger targeting range of CRISPR/Cas9 relative to RNAi extends to the whole genome and offers the opportunity to probe structure/function relationships beyond the transcriptome. As well, the potential exists for Cas9-driven cleavage events to yield not only loss-of-function but also gain-of-function and dominant-negative, alleles—thus extending the mutational "depth" beyond the straight suppression possible with RNAi. Whereas somatic cell genetics provided stunning insights into gene organization and regulation in the 1970s and 1980s (Caskey, Robbins, North Atlantic Treaty Organization, & Scientific Affairs Division, 1982), the remarkable progress that has been made in applying CRISPR/Cas9 to genome engineering since 2013 and the potential it holds for genetic analysis of almost any cell type at an unprecedented scale would suggest an up-and-coming rebirth of this discipline. It will be exciting to participate in this new adventure as CRISPR/Cas9 is used to uncover novel genome functionalities.

REFERENCES

Barde, I., Salmon, P., & Trono, D. (2001). Production and titration of lentiviral vectors. *Current protocols in neuroscience*. New York, NY: John Wiley & Sons, Inc.

Caskey, C. T., Robbins, D. C., North Atlantic Treaty Organization, & Scientific Affairs Division. (1982). *Somatic cell genetics*. New York: Plenum Press, published in cooperation with NATO Scientific Affairs Division.

Certo, M. T., Ryu, B. Y., Annis, J. E., Garibov, M., Jarjour, J., Rawlings, D. J., et al. (2011). Tracking genome engineering outcome at individual DNA breakpoints. *Nature Methods*, *8*(8), 671–676. http://dx.doi.org/10.1038/nmeth.1648.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479–1491. http://dx.doi.org/10.1016/j.cell.2013.12.001.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823. http://dx.doi.org/10.1126/science.1231143, science.1231143 [pii].

Ellison, A. R., & Bishop, J. O. (1996). Initiation of herpes simplex virus thymidine kinase polypeptides. *Nucleic Acids Research*, *24*(11), 2073–2079.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., et al. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*, *31*(9), 822–826. http://dx.doi.org/10.1038/nbt.2623, nbt.2623 [pii].

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, *32*(3), 279–284. http://dx.doi.org/10.1038/nbt.2808, nbt.2808 [pii].

Green, M. R., & Sambrook, J. (2012). *Molecular cloning: A laboratory manual* (4th ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–832. http://dx.doi.org/10.1038/nbt.2647, nbt.2647 [pii].

Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, *31*(3), 233–239. http://dx.doi.org/10.1038/nbt.2508, nbt.2508 [pii].

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. http://dx.doi.org/10.1126/science.1225829, science.1225829 [pii].

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife*, *2*, e00471. http://dx.doi.org/10.7554/eLife.00471 00471 [pii].

Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D., & Yusa, K. (2013). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnology*, *32*, 267–273. http://dx.doi.org/10.1038/nbt.2800.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., & Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*, *32*, 677–683. http://dx.doi.org/10.1038/nbt.2916, nbt.2916 [pii].

Kutner, R. H., Zhang, X.-Y., & Reiser, J. (2009). Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nature Protocols*, *4*(4), 495–505. http://dx.doi.org/10.1038/nprot.2009.22.

Lin, B. R., & Natarajan, V. (2012). Negative regulation of human U6 snRNA promoter by p38 kinase through Oct-1. *Gene*, *497*(2), 200–207. http://dx.doi.org/10.1016/j.gene.2012.01.041.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., et al. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology*, *31*(9), 833–838. http://dx.doi.org/10.1038/nbt.2675, nbt.2675 [pii].

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., et al. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826. http://dx.doi.org/10.1126/science.1232033 science.1232033 [pii].

Malina, A., Mills, J. R., Cencic, R., Yan, Y., Fraser, J., Schippers, L. M., et al. (2013). Repurposing CRISPR/Cas9 for in situ functional assays. *Genes & Development*, *27*(23), 2602–2614. http://dx.doi.org/10.1101/gad.227132.113.

Nielsen, S., Yuzenkova, Y., & Zenkin, N. (2013). Mechanism of eukaryotic RNA polymerase III transcription termination. *Science*, *340*(6140), 1577–1580. http://dx.doi.org/10.1126/science.1237934.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). Crystal structure of cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935–949. http://dx.doi.org/10.1016/j.cell.2014.02.001.

Orioli, A., Pascali, C., Quartararo, J., Diebel, K. W., Praz, V., Romascano, D., et al. (2011). Widespread occurrence of non-canonical transcription termination by human RNA polymerase III. *Nucleic Acids Research*, *39*(13), 5499–5512. http://dx.doi.org/10.1093/nar/gkr074.

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, *31*(9), 839–843. http://dx.doi.org/10.1038/nbt.2673, nbt.2673 [pii].

Reyon, D., Tsai, S. Q., Khayter, C., Foden, J. A., Sander, J. D., & Joung, J. K. (2012). FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology*, *30*(5), 460–465. http://dx.doi.org/10.1038/nbt.2170.

Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10098–10103. http://dx.doi.org/10.1073/pnas.1104144108, 1104144108 [pii].

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, *343*(6166), 84–87. http://dx.doi.org/10.1126/science.1247005.

Swift, S., Lorens, J., Achacoso, P., & Nolan, G. P. (2001). Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. *Current protocols in immunology*. New York, NY: John Wiley & Sons, Inc.

Szymczak-Workman, A. L., Vignali, K. M., & Vignali, D. A. (2012). Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harbor Protocols*, *2012*(2), 199–204. http://dx.doi.org/10.1101/pdb.ip067876.

Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, *343*(6166), 80–84. http://dx.doi.org/10.1126/science.1246981.

Wu, X., Scott, D. A., Kriz, A. J., Chiu, A. C., Hsu, P. D., Dadon, D. B., et al. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, *32*, 670–676. http://dx.doi.org/10.1038/nbt.2889, nbt.2889 [pii].

Yin, H., Xue, W., Chen, S., Bogorad, R. L., Benedetti, E., Grompe, M., et al. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature Biotechnology*, *32*, 551–553. http://dx.doi.org/10.1038/nbt.2884, nbt.2884 [pii].

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., et al. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, *509*(7501), 487–491. http://dx.doi.org/10.1038/nature13166, nature13166 [pii].